2 Probability Theory

2.1 INTRODUCTION

In this chapter, we present a theory of probability that is based on *three axioms* and is consistent with our intuition about the probability of an event. This will require a review of some background material on set theory, which will lead to definitions of an *event* and a *field* that are needed so that probabilities can be assigned and computed. Perhaps the simplest and most widely used example of a random experiment is the toss of a single coin. Although relatively trivial, we use this experiment often to give definitions and describe properties of random events. Since there are only two outcomes in this experiment, heads (H) and tails (T), it is also a model for a digital communication system which employs a binary *alphabet* $\{0, 1\}$ (outcomes $\{-1, 1\}$ are also used when symmetry about zero is desired). Even though initially the coin-toss experiment appears to be lacking much structure, it is actually quite useful and relevant for many applications.

Another experiment that will be used to demonstrate various properties of probability is the toss of a single die, which has more complexity than the coin-toss experiment because there are multiple outcomes: $\{1, 2, 3, 4, 5, 6\}$. Although any experiment with several outcomes would be sufficient for our purposes, we use the die-toss experiment because it is familiar to most readers. This example also has relevance in communications, as a model for pulse amplitude modulation (PAM) where *symbols* are selected from a finite alphabet containing 4, 8, 16, and so on, elements for transmission across a channel.

With these two simple experiments, we can describe the important concepts needed to construct a probability space for experiments with a *finite* number of outcomes. With an understanding of these examples, the probability model can be extended to experiments with an *infinite* number of outcomes. There are three basic types of random experiments as determined by the number of outcomes: (i) a finite number of outcomes (such as a single coin toss), (ii) an infinite but *countable* number of outcomes (such as tossing a single coin ad infinitum), and (iii) a continuum (uncountable number) of outcomes (such as temperature measurements). More complicated experiments are combinations of these three types of outcomes.

It is relatively straightforward to assign probabilities to events for experiments in category (i). Extending those results to experiments in category (ii) requires some work, but is also straightforward because the outcomes are countable: there is a one-to-one mapping of outcomes to the natural numbers $\mathcal{N} = \{1, 2, \ldots\}$. Experiments in category (iii) are the most difficult to quantify because there is a continuum of outcomes: for any two values $a < b$, there exists $\epsilon > 0$ such that $a < a + \epsilon < b$. Thus, it is not immediately obvious how to assign probabilities to an uncountable number of outcomes; in fact, it turns out that the probability of a single outcome in a continuous experiment is zero. Events in continuous experiments require special consideration when specifying the probability space.

As a preview, we summarize how probabilities are generally computed for each of the three types of experiments:

- *Finite*. Probabilities are computed by *counting* the outcomes comprising an event of interest, relative to the total number of outcomes in the sample space Ω. Basic rules of *combinatorics* are used for finite problems.

Probability, Random Variables, and Random Processes: Theory and Signal Processing Applications, First Edition. John J. Shynk.

TABLE 2.1 Basic Types of Random Experiments with Examples

Outcomes	Finite	Infinite
Discrete	Tossing two coins (*finite*)	Repeated tosses of a single coin (*countably infinite*, denumerable)
Continuous	None	Temperature (*uncountable*, nondenumerable)

- *Countably infinite*. Probabilities are computed by *summing* over a discrete probability mass function that characterizes the experiment. Techniques for computing finite and *infinite sums* are used for countably infinite problems.
- *Uncountable*. Probabilities are computed by *integrating* over a continuous probability density function that characterizes the experiment. Techniques from *calculus* are used for uncountable problems.

Finite experiments can also be solved using techniques for countably infinite problems, but it is often easier to simply count outcomes as mentioned above. Countably infinite and uncountable experiments problems are not easily examined without a more rigorous characterization, as provided by the *probability mass function* (pmf) and the *probability density function* (pdf). As shown later, it is also possible to handle countably infinite experiments using methods from calculus (via the Dirac delta function), but the results essentially reduce to evaluating finite and infinite sums.

Examples for each of the three cases are summarized in Table 2.1. It is obvious that by tossing two coins there are only four outcomes: $\{HH, TT, HT, TH\}$. If the probability of at least one H is of interest, we simply count the number of such outcomes (which is three) and divide that by the total number of outcomes. Thus, the probability of at least one H is 0.75; it is denoted by $P(\text{at least one } H) = P(HH \text{ or } HT \text{ or } TH) = 0.75$. For the countably infinite example in the table, we might be interested in the probability that H first appears after the second toss. Clearly, this is more difficult to compute than the previous finite-outcome example: we need to consider all *sequences* of the form $TTHT\cdots$, $TTHH\cdots$, $TTTH\cdots$, and so on. As shown later, the pmf allows for a straightforward probability computation, which is easier than attempting to directly count all outcomes of interest. Likewise for the uncountable (continuous) case, where the pdf is used to compute probabilities via an integration over the outcomes of interest, which are described by *intervals* on $\mathcal{R}$.

Although we use the term *experiment* to describe how random outcomes are generated, in real applications the notion of experiment is artificial because events "just occur" (such as temperature fluctuations, radioactive decay, and so on). However, we use the terms experiment and *trial* not only for a *synthetic* experiment, such as tossing a die or using a computer program to generate a random number, but also for *natural* events. In the latter case, the experiment can be viewed as a model of the underlying mechanisms for such events.

A probabilistic modeling of events in the physical world is advantageous because most phenomena are too complicated to accurately represent using physical models. For example as mentioned in Chapter 1, the toss of a single coin could be modeled using many factors, such as the height from which it is tossed, the velocity and angle of the toss, and the type of surface upon which it lands, and so on. Of course, such a model for even this simple experiment is complicated and undoubtedly incomplete. By using a probabilistic model, we bypass the need for an accurate physical model, and can make predictions about the occurrence of an event with a high degree of accuracy, and which is consistent with our intuition about randomness (the *frequency* of outcomes as mentioned in Chapter 1). A probabilistic model is a powerful representation that allows us to "say something" about events that arise due to the complicated interactions of many underlying physical mechanisms.

2.2 SETS AND SAMPLE SPACES

We begin with basic definitions of sets and sample spaces, which should be a review for most readers.

Definition: Set A *set* is a collection of objects or numbers that represent those objects. These objects are called *elements* or *points* of the set.

Although it is possible to accurately describe a set using the original objects or outcomes in many experiments, it is often cumbersome, especially for a large number of elements. Instead, we represent the elements of a set using numbers—either integers, real numbers, or complex numbers—because it is more convenient mathematically, and is actually necessary when defining *random variables* in Chapter 3. For example, in experiments involving the toss of a single coin, it will be convenient to assign numbers such as $1 \equiv H$ and $0 \equiv T$ to represent the two outcomes.

Sets will be denoted by uppercase letters, usually at the beginning of the Latin alphabet, as illustrated in the following examples.

Example 2.1. Set $A = \{2, 4, 6, 8\}$ consists of four integers, and set $B = \{\ldots, -1, 0, 1, 2, \ldots\}$ consists of all integers.

The set of all integers is represented by $\mathcal{Z}$ and positive integers by $\mathcal{Z}^+$, which *includes* zero.

Example 2.2. Set $C = (1, 5)$ consists of all real numbers between 1 and 5 *excluding* 1 and 5, whereas set $D = [1, 5]$ includes the end points.

The set of all real numbers is represented by $\mathcal{R}$ and positive real numbers by $\mathcal{R}^+$, which *includes* zero. We often use the symbols $\pm\infty$ which technically are not real numbers. Although complex quantities are not used extensively in this book, they do appear in the applications part; the symbol for all complex numbers is $\mathcal{C}$.

Element a in set A is written as $a \in A$; element b that is not in set A is written as $b \notin A$. The type of set can be characterized according to the number of objects.

Definition: Cardinality The *cardinality* of a set is the number of elements in that set. The cardinality of set E is denoted by $|E|$.

A countable set can have either a finite or an infinite number of elements; the former is called *finite cardinality* $< |\mathcal{N}|$), and the latter is *countably infinite* or *denumerable* (cardinality $= |\mathcal{N}|$). The elements of a countably infinite are in one-to-one correspondence with the natural numbers $\mathcal{N} \triangleq \{1, 2, \ldots\}$ (note that $\mathcal{N}$ does *not* include zero).

Example 2.3. $\mathcal{Z}$ and $\mathcal{Z}^+$ have the same cardinality as $\mathcal{N}$ and thus are countably infinite. The rational numbers, denoted by $\mathcal{Q}$ and having the form a/b with $a, b \in \mathcal{Z}$ $(b \neq 0)$ also have cardinality $|\mathcal{N}|$.

Example 2.4. The irrational numbers have a cardinality $> |\mathcal{N}|$ and are uncountable. Likewise, $\mathcal{R}$ and any interval on $\mathcal{R}$ have a cardinality $> |\mathcal{N}|$, and both are uncountable. "Almost all" irrational numbers are transcendental, of which π and e are perhaps the most well known of the set.

Set A in Example 2.1 is finite, whereas set B is countably infinite. If an infinite set is not countable, then it is *uncountable* or *nondenumerable* with cardinality $> |\mathcal{N}|$. Set C in Example 2.2 is uncountable. Countable sets are also called *discrete* sets, while uncountable sets are *continuous*.

These set distinctions are useful later when developing a probability space for events of an experiment. It is necessary that we precisely define events of an experiment and the corresponding *algebra* of operations which allow for a consistent probability measure. Figure 2.1 summarizes the three basic types of sets with simple examples. An example of a *mixed* set consisting of discrete and continuous components is also included. Although a mixed set is by definition uncountable (the fact that it has a discrete component does not change the fact that the continuous part is uncountable), it will be convenient to identify it as having a finite or countably infinite component when discussing random variables in Chapter 3. In the example shown in Figure 2.1, note that the union operator $\cup$ is used to describe the entire set; this set operation will be covered later.

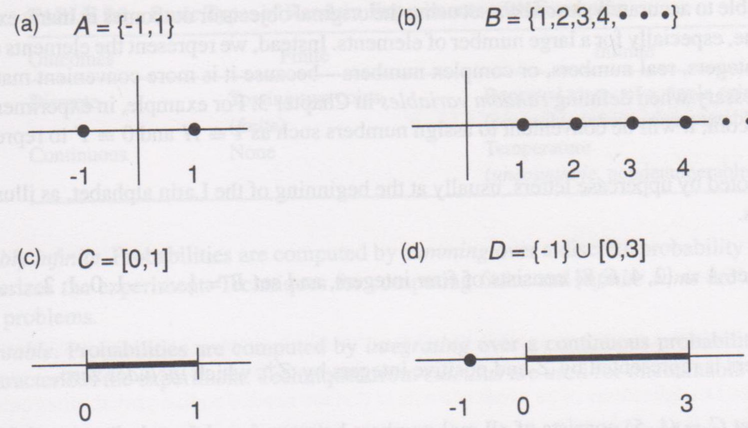

FIGURE 2.1 Examples of the three basic types of sets and a mixed set. (a) Finite. (b) Countably infinite. (c) Uncountable (an interval). (d) Mixed (an interval and one point).

Regarding set notation, we generally use braces $\{\cdot\}$ to denote a discrete set (either finite or countably infinite), such as in Example 2.1. However, it is also convenient to use braces when a *function* describes a continuous set, as demonstrated by the following example.

Example 2.5. Discrete set: $A = \{x^3 - 1 : x = 0, 2, 4\} = \{-1, 7, 63\}$. Continuous set: $B = \{x : x \geq -1\} = [-1, \infty)$. A set defined in terms of other sets: $C = \{x : x \in A \text{ or } x \in B\} = A \cup B$.

The colon (:) in this context should be read as "such that" (s.t.). The last case describes the union operation "or" ($\cup$): C is the set of all x such that x is in set A or B or *both*. Parentheses $(\cdot)$ and brackets $[\cdot]$ are also used to describe uncountable sets, such as those in Examples 2.2 and 2.5. The various types of *intervals* for uncountable sets are defined next.

Definition: Intervals and Singletons The *open interval* (a, b) is the set containing all real numbers between a and b (with $a < b$), but excluding the boundary points a and b. The *closed interval* $[a, b]$ is the set containing all real numbers between a and b, including the boundary points a and b. The *semi-open intervals* $[a, b)$ and $(a, b]$ are obtained from $[a, b]$ by excluding one of the boundary points: b is excluded in $[a, b)$ and a is excluded in $(a, b]$. The boundary points are called ***singletons***, as is any individual point in $\mathcal{R}$.

We can also define intervals using the following brace notation (as was done in Example 2.5):

$$(a, b) \triangleq \{x : a < x < b\}, \quad (a, b] \triangleq \{x : a < x \leq b\},$$
$$[a, b) \triangleq \{x : a \leq x < b\}, \quad [a, b] \triangleq \{x : a \leq x \leq b\}. \tag{2.1}$$

Since an interval is a continuous set, the following mathematical description for an open interval can be formulated. If $x \in (a, b)$, then there exists $\epsilon > 0$ such that $x + \epsilon \in (a, b)$ and $x - \epsilon \in (a, b)$. Regardless of how close x is to a or b in an *open interval*, we can always find numbers between $x \in (a, b)$ and the points a and b. Of course, this property does not apply to semi-open and closed intervals because x could be chosen as one of the boundary points.

In order to specify events and their probabilities, we define the basic components of an experiment.

Definition: Sample Space Ω The *sample space* Ω is the set of all possible outcomes in an experiment. It is also called the *universe* or the *universal set*. An element of Ω is denoted by ω and may include a subscript such as ω_n.

Example 2.6. $\Omega = \{0, 1\}$ for the two possible symbols used in a binary digital communication system. This also corresponds to the sample space for a single toss of a coin with $H \equiv 1$ and $T \equiv 0$.

Example 2.7. $\Omega = [0, \infty) = \mathcal{R}^+$ for the infinity of outcomes in an experiment using the set of nonnegative real numbers. For many experiments, $\Omega = (-\infty, \infty)$ which is the set of all real numbers $\mathcal{R}$. Note that we always use an open interval to exclude $\pm\infty$ because, as mentioned earlier, they are not real numbers.

For a group of sets in the universe Ω, it is possible to define some fundamental set relationships.

Definition: Subset $A \subset B$ A is a *subset* of B if all elements of A are in B, whereas B may have elements that are not in A. The notation $A \subseteq B$ allows for the possibility that A and B are identical sets.

Example 2.8. If $A = [0, 5]$ and $B = [0, 10)$, then $A \subset B$. The semi-open interval $(a, b]$ is a subset of the closed interval $[a, b]$. Clearly, any set is a subset of itself: $A \subset A$. If all elements of B are also in A, then the two sets are equal.

Definition: Equality $A = B$ Sets A and B are *equal* if they have exactly the same elements such that $A \subset B$ and $B \subset A$.

Since the sample space Ω contains all outcomes for an experiment, we can describe elements that exist outside of a particular set.

Definition: Complement A^c The *complement* of A is the set of elements in Ω that are not in A. Thus, $a \notin A$ implies that $a \in A^c$. Using the brace notation: $A^c = \{x : x \notin A\}$.

Example 2.9. If $\Omega = [0, 100]$ and $A = [0, 25)$, then $A^c = [25, 100]$. If $\Omega = \mathcal{R}$ and $B = (0, \infty)$, then $B^c = (-\infty, 0]$.

The complement of A is also denoted by $\bar{A}$, but we use A^c because the bar notation is used later for sample moments (such as the sample mean). Note that Ω *must* be defined in order for A^c to have meaning. We conclude this section with two important sets.

Definition: Empty Set ϕ The *empty set* is the set containing no elements. It is also sometimes called the null set. From the previous definitions, it is clear that $\phi = \Omega^c$.

Sometimes the empty set is denoted by $\{\}$. The empty set refers to the event "nothing occurs." As discussed later, the empty set is needed when defining an *algebra* of operations for sets.

Definition: Power Set $\mathcal{P}(\Omega)$ The *power set* contains all possible subsets of the sample space Ω. It includes the empty set ϕ as well as Ω. The power set can also be defined for any subset $E \subset \Omega$; it is denoted by $\mathcal{P}(E)$.

Example 2.10. If $\Omega = \{2, 4, 6\}$, then $\mathcal{P}(\Omega) = \{\phi, \Omega, 2, 4, 6, \{2, 4\}, \{2, 6\}, \{4, 6\}\}$ has eight elements. The power set of $E = \{2, 4\} \subset \Omega$ is $\mathcal{P}(E) = \{\phi, E, 2, 4, \{2, 4\}\}$. The power set of a countably infinite set (such as $\mathcal{N}$) is not easily visualized, but it is uncountably infinite and has the same cardinality as $\mathcal{R}$.

If E is finite, then $|\mathcal{P}(E)|$ is necessarily finite.

Proposition 2.1. *The power set of E consisting of N elements has cardinality* $|\mathcal{P}(E)| = 2^N$.

Proof. This is proved using the binomial theorem in Appendix E which states that

$$(x+y)^N = \sum_{n=0}^{N} \binom{N}{n} x^{N-n} y^n, \tag{2.2}$$

where

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \tag{2.3}$$

is a binomial coefficient (discussed later in connection with probabilities for finite sample spaces). A binomial coefficient is the number of subsets (combinations) of size n chosen from N elements. Summing over n to count all subsets and letting $x = y = 1$ in (2.2) yields the number of elements in $\mathcal{P}(E)$:

$$\sum_{n=0}^{N} \binom{N}{n} = (1+1)^N = 2^N, \tag{2.4}$$

which completes the proof. □

With these basic definitions involving sets, we can consider operations on sets.

2.3 SET OPERATIONS

Most problems of interest in probability involve operations on events. The two basic set operations are union and intersection.

Definition: Union $A \cup B$ The *union* of two sets is the set of all elements in A or B or both:

$$A \cup B = \{x : x \in A \text{ or } x \in B\}. \tag{2.5}$$

The sum notation $A + B$ is sometimes used to denote the union operation.

Definition: Intersection $A \cap B$ The *intersection* of two sets is the set of elements common to A and B:

$$A \cap B = \{x : x \in A \text{ and } x \in B\}. \tag{2.6}$$

The product notation AB is sometimes used to denote the intersection operation.

Example 2.11. If $A = \{1, 2, 3\}$ and $B = \{3, 4\}$, then $A \cup B = \{1, 2, 3, 4\}$ and $A \cap B = \{3\}$. If $E = [2, \infty)$ and $F = (-\infty, 5]$, then $E \cup F = \mathcal{R}$ and $E \cap F = [2, 5]$. If $\Omega = \mathcal{R}$, $A = (a, b)$, and $B = [a, b]$, then $AB = A$, $A \cup B = B$, $A^c B = \{a, b\}$ is finite, and $AB^c = \phi$.

The union and intersection operations can be extended to any number of sets, including an infinite collection. Examples of the notation used for N sets $\{A_n\}$ are

$$B_1 = \bigcup_{n=1}^{N} A_n, \qquad B_2 = \bigcap_{n=1}^{N} A_n. \tag{2.7}$$

For convenience, we generally use the shorthand notation AB for intersection (though this may not always be possible), but will retain the conventional notation $A \cup B$ for union because $A + B$ could be confusing later when the probabilities of events are added.

It is straightforward to show that union and intersection satisfy the following three basic laws:

- *Commutative:*

$$A \cup B = B \cup A, \qquad AB = BA. \tag{2.8}$$

- *Associative:*

$$(A \cup B) \cup C = A \cup (B \cup C), \qquad (AB)C = A(BC). \tag{2.9}$$

- *Distributive:*

$$A \cup (BC) = (A \cup B)(A \cup C), \qquad A(B \cup C) = (AB) \cup (AC). \tag{2.10}$$

These operations also satisfy De Morgan's laws which combine intersection, union, and complement as follows:

- *De Morgan's laws:*

$$(A \cup B)^c = A^c B^c, \qquad (AB)^c = A^c \cup B^c. \tag{2.11}$$

These laws are all easily verified using a Venn diagram, which is described later, and they are readily extended to any number of sets.

Next, we illustrate how intervals behave under union and intersection operations, which are not obvious for *infinite* unions and intersections arising in continuous sample spaces. For appropriate values of $\{a_n\}$, $\{b_n\}$, c, and d, the union and intersection of intervals on the real line $\mathcal{R}$ have the following properties:

- A *finite* union or intersection of *closed* intervals yields a *closed* interval:

$$\bigcup_{n=1}^{N} [a_n, b_n] = [c, d], \qquad \bigcap_{n=1}^{N} [a_n, b_n] = [c, d]. \tag{2.12}$$

- A *finite* union or intersection of *open* intervals yields an *open* interval:

$$\bigcup_{n=1}^{N} (a_n, b_n) = (c, d), \qquad \bigcap_{n=1}^{N} (a_n, b_n) = (c, d). \tag{2.13}$$

- A *countably infinite union* of *open* intervals yields an *open* interval:

$$\bigcup_{n=1}^{\infty} (a_n, b_n) = (c, d). \tag{2.14}$$

- A *countably infinite intersection* of *closed* intervals yields a *closed* interval:

$$\bigcap_{n=1}^{\infty} [a_n, b_n] = [c, d]. \tag{2.15}$$

The first two properties involving a finite number of open or closed sets are obvious. It is relatively straightforward to verify the third and fourth properties, especially after we discuss the next two properties which are not obvious. The operations in the four cases above all give the same type of interval as the original intervals.

- A *countably infinite union* of *closed* intervals can yield *any* type of interval (open, semi-open, or closed):

$$\bigcup_{n=1}^{\infty}[a_n, b_n] = (c, d) \text{ or } (c, d] \text{ or } [c, d) \text{ or } [c, d]. \tag{2.16}$$

- A *countably infinite intersection* of *open* intervals can yield *any* type of interval (open, semi-open, or closed):

$$\bigcap_{n=1}^{\infty}(a_n, b_n) = [c, d] \text{ or } (c, d] \text{ or } [c, d) \text{ or } (c, d). \tag{2.17}$$

These last two properties are interesting because the set operations can result in a type of interval that is different from the component intervals. Nothing in general can be said about the *infinite union* of *closed* intervals, nor about the *infinite intersection* of *open* intervals. The following two examples illustrate this behavior.

Example 2.12. If $(a, b_n) = (-1, 1/n)$, then $\bigcap_{n=1}^{\infty}(a, b_n) = (-1, 0]$. Observe that all the open intervals include zero, and thus their intersection—even an infinite number of them—must also include zero, so the intersection yields a semi-closed interval.

Example 2.13. If $[a_n, b] = [1/n, 1]$, then $\bigcup_{n=1}^{\infty}[a_n, b] = (0, 1]$. It is clear that since none of the sets include zero, their union can never contain zero, even an infinite number of unions. The resulting interval must be semi-open.

Variations of these examples can be devised to illustrate the other three types of intervals in (2.16) and (2.17). For convenience, Table 2.2 summarizes the possible results for set operations involving intervals. An entry in the table with multiple intervals means that the set operation *could* take on any one of those forms (though not necessarily), depending on how the initial interval in the first column is defined. We mention these results because an algebra of sets must be closed under the union and intersection operations, and thus a σ-field (defined later for $\mathcal{R}$) cannot consist solely of one type of interval.

The intersection operation is useful for defining the following important property of sets.

Definition: Mutually Exclusive Sets A and B are *mutually exclusive* if they have no elements in common such that $AB = \phi$. Mutually exclusive sets are also called *disjoint*.

This definition can be extended to any number of sets $\{A_n\}$ such that the intersection of any two sets is disjoint (called pairwise disjoint): $A_n A_m = \phi$ for all $n \neq m$. Moreover, a sample space can be divided into a collection

TABLE 2.2 Set Operations of Intervals

Type of Interval	Finite Union	Finite Intersection	Countably Infinite Union	Countably Infinite Intersection
$[a_n, b_n]$	$[c, d]$	$[c, d]$	$(c, d), (c, d], [c, d), [c, d]$	$[c, d]$
$[a_n, b_n)$	$[c, d)$	$[c, d)$	$(c, d), [c, d)$	$[c, d), [c, d]$
$(a_n, b_n]$	$(c, d]$	$(c, d]$	$(c, d), (c, d]$	$(c, d], [c, d]$
(a_n, b_n)	(c, d)	(c, d)	(c, d)	$(c, d), (c, d], [c, d), [c, d]$

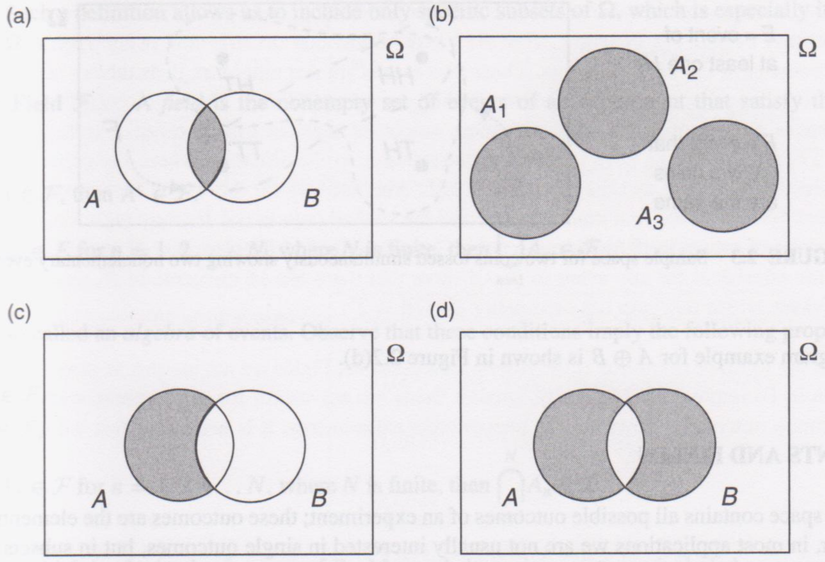

FIGURE 2.2 Set operations (denoted by the shaded regions) are visualized using Venn diagrams. (a) $A \cap B$. (b) Partition: $A_1 \cup A_2 \cup A_3 = \Omega$. (c) Difference $A - B$. (d) Exclusive or $A \oplus B$.

of disjoint sets that together contain all elements of Ω. Such a collection of sets is called a partition of the sample space.

Definition: Partition Disjoint sets $\{A_n\}$ form a *partition* of Ω if $\bigcup_n A_n = \Omega$.

Partitions of a sample space are not unique, and the number of possible partitions is finite only if Ω is finite. If the sets $\{A_n\}$ are not disjoint but we still have $\cup_n A_n = \Omega$, then $\{A_n\}$ are *collectively exhaustive*.

It is convenient at this point to utilize a *Venn diagram* to illustrate various set relationships. A Venn diagram is a simple pictorial representation of sets, often using circles to represent sets within a sample space that illustrate how they are related by their degree of overlap. Figure 2.2(a) shows a Venn diagram for $A \cap B$. Note that $A \cap B = \phi$ if A and B are disjoint, and $A \cup B = \Omega$ because in this example there are no elements outside A and B. Figure 2.2(b) shows a partition of a sample space into three disjoint sets $\{A_1, A_2, A_3\}$. The circles are nonoverlapping, and all points of the sample space are contained in the union of these three sets: $A_1 \cup A_2 \cup A_3 = \Omega$.

We conclude this section with two additional set operations that are frequently useful.

Definition: Difference $A - B$ The *difference* of sets A and B is the set of elements in A excluding those also in B. It can be written explicitly using previous set definitions as $A - B = AB^c$. The notation $A \setminus B$ is sometimes used to denote the difference of two sets.

A Venn diagram illustrating the difference of sets A and B is shown in Figure 2.2(c). Observe that if A and B are disjoint, then $A - B = A$ and $B - A = B$. The difference operator does not satisfy the commutative law: $A - B \neq B - A$ (except trivially when $B = A$ such that $A - B = \phi$). The difference of two sets is not a symmetric operation. It is easy to verify $A - B = AB^c$ using a Venn diagram.

Definition: Exclusive Or $A \oplus B$ The *exclusive or* operation yields the set of elements in A or B but *not both*: $A \oplus B = A \cup B - AB = (A - B) \cup (B - A)$. Because of this last formulation, the exclusive or operation is also called the *symmetric difference*.

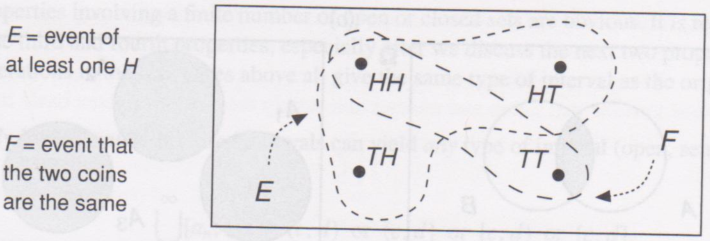

FIGURE 2.3 Sample space for two coins tossed simultaneously showing two nonelementary events.

A Venn diagram example for $A \oplus B$ is shown in Figure 2.2(d).

2.4 EVENTS AND FIELDS

The sample space contains all possible outcomes of an experiment; these outcomes are the elements (points) of Ω. However, in most applications we are not usually interested in single outcomes, but in subsets of Ω called events.

Definition: Event *Event E* of an experiment is a subset of the sample space Ω whose elements share a common feature.

Example 2.14. The experiment where two coins are tossed simultaneously has the four outcomes shown in Figure 2.3 which are the four elements of Ω. Consider the event of at least one H given by $E = \{HH, HT, TH\}$. Thus, E is not a specific outcome, but is instead a *feature* of three of the four possible outcomes. Obviously, E is a subset of Ω.

Example 2.15. For the experiment of tossing a single die, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. Subset $E = \{1, 3, 5\}$ is the event that the die shows an odd value.

Example 2.16. In an experiment where two dice are tossed, the event that they show the same value is the subset $A = \{\{1, 1\}, \{2, 2\}, \{3, 3\}, \{4, 4\}, \{5, 5\}, \{6, 6\}\}$.

Example 2.17. For $\Omega = [0, 1]$, $A = [0, 1/2]$ and $B = [0, 1/4] \cup [5/8, 7/8]$ are examples of events.

Note that an element $\omega_n \in \Omega$ corresponding to a single outcome is a particular type of event.

Definition: Elementary Event An *elementary event* is a subset of Ω consisting of one element, and is the same as a single outcome.

Sometimes an elementary event is also called an atomic event, but we reserve the name *atom* for a specific component of the σ-field defined later.

Example 2.18. Subset $F = \{1\}$ in Example 2.15 is an elementary event, as is the subset $B = \{2, 2\}$ in Example 2.16.

Since our goal is to assign probabilities to events of an experiment, it is important that events be defined to satisfy *algebraic rules* that allow for consistency in the probability space. In order to be rigorous about operations involving events, we define a *field* as a collection of events that satisfy certain algebraic conditions. Such a definition allows us to include only specific subsets of Ω, which is especially important for continuous Ω.

Definition: Field $\mathcal{F}$ A *field* is the nonempty set of events of an experiment that satisfy the following conditions:

(i) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

(ii) If $A_n \in \mathcal{F}$ for $n = 1, 2, \ldots, N$, where N is finite, then $\bigcup_{n=1}^{N} A_n \in \mathcal{F}$.

A field is also called an *algebra* of events. Observe that these conditions imply the following properties:

(iii) $\Omega \in \mathcal{F}$.

(iv) $\phi \in \mathcal{F}$.

(v) If $A_n \in \mathcal{F}$ for $n = 1, 2, \ldots, N$, where N is finite, then $\bigcap_{n=1}^{N} A_n \in \mathcal{F}$.

Property (v), which is obtained using one of De Morgan's laws, is sometimes included as a condition in the definition of a field, but it need not be mentioned explicitly because (i) and (ii) are sufficient.

Example 2.19. The simplest (but trivial) field is $\mathcal{F} = \{\phi, \Omega\}$.

Example 2.20. For the experiment of a single coin toss with $\Omega = \{H, T\}$, consider the nontrivial field $\mathcal{F} = \{\phi, \Omega, H, T\}$. The properties above are readily verified: (i) $H \in \mathcal{F} \Longrightarrow T \in \mathcal{F}$ (and vice versa), (ii)–(iii) $H \in \mathcal{F}, T \in \mathcal{F} \Longrightarrow H \cup T \equiv \Omega \in \mathcal{F}$, and (iv)–(v) $H \in \mathcal{F}, T \in \mathcal{F} \Longrightarrow H \cap T \equiv \phi \in \mathcal{F}$. Note that $\mathcal{F}$ is, in fact, the power set $\mathcal{P}(\Omega)$ for this experiment, and no other fields are possible, except the trivial field in Example 2.19.

In some experiments, we might first consider a set of events that do not quite comprise a field, and then include additional elements to *generate* a field. This procedure is illustrated in Example 2.21.

Example 2.21. Suppose that $\Omega = [1, \infty)$ and consider $\mathcal{F} = \{\phi, \Omega, \{[1, a)\}\}$ for every $a > 1$, that is, $\mathcal{F}$ contains all semi-open intervals of the form $[1, a)$. Obviously, $\mathcal{F}$ is not a field because $[1, a)^c = [a, \infty)$ for any $a > 1$ is not in $\mathcal{F}$. Expand the set to be $\mathcal{F} = \{\phi, \Omega, \{[1, a)\}, \{[b, \infty)\}\}$ with $b > 1$. Again, this is not a field because $[1, a) \cap [b, \infty) = [b, a)$ for any $b < a$ is not in $\mathcal{F}$. Expand the set further to be $\mathcal{F} = \{\phi, \Omega, \{[1, a)\}, \{[b, \infty)\}, \{[c, d)\}\}$ for all $a, b, c, d > 1$. We now have a field: all *finite* unions of the elements are events in $\mathcal{F}$, as are all complements of the elements. Later, we demonstrate that $\mathcal{F}$ is *not* a σ-field which is defined for an *infinity* of unions, and is described after the next set of examples.

Example 2.22. For the experiment of tossing two coins, recall that $\Omega = \{HH, TT, HT, TH\}$. The elements of the power set $\mathcal{P}(\Omega)$ are summarized in Table 2.3. Properties (i) and (ii) are easily verified for this case. From this example, it is useful to summarize some characteristics of an event. The four outcomes in the experiment

TABLE 2.3 Elements of Field $\mathcal{F} = \mathcal{P}(\Omega)$ for Example 2.22

Elementary events (the four outcomes):	$\{HH\}, \{TT\}, \{HT\}, \{TH\}$
Events defined by two outcomes:	$\{HH, TT\}, \{HH, HT\}, \{HH, TH\}, \{TT, HT\}, \{TT, TH\}, \{HT, TH\}$
Events defined by three outcomes:	$\{HH, TT, HT\}, \{HH, TT, TH\}, \{HH, HT, TH\}, \{TT, HT, TH\}$
ϕ and Ω:	$\phi = \Omega^c, \Omega = \{HH, TT, HT, TH\}$

are given by the terms in the first row of Table 2.3. These are the elementary events; all other entries in the table are not elementary events. For example, $\{HH, HT\}$ is the event that the first coin is H; $\{HH, TT\}$ is the event that the two coins have the same outcome. These two events are *not* outcomes themselves; instead, they refer to those outcomes that share a common feature. Although event $\{HH, HT, TH\}$ corresponds to those outcomes such that there is at least one H, events with three outcomes together can be more complicated to describe. Event $\{HH, TT, HT\}$ is not so easy to state; it corresponds to the situation where the coins have the same outcome *or* the first coin is H and the second coin is T. Alternatively for this event, we can say that it refers to those outcomes for which the first coin is not T *and* the second coin is not H (thus excluding TH). Since we require all unions of events in $\mathcal{F}$ to be in the field, Ω must be in $\mathcal{F}$. Likewise, $\phi \in \mathcal{F}$ so that complement and intersection are consistent for any event in $\mathcal{F}$. Observe that there are 16 elements in $\mathcal{F}$, as expected because $N = 4$; the power set for this case has cardinality $|\mathcal{P}(\Omega)| = 2^4$.

The power set $\mathcal{P}(\Omega)$ by definition is the largest field possible. Fields are not unique, as mentioned previously and as shown in Example 2.23. Generally, smaller fields are not useful for experiments with finite and even countably infinite outcomes. However, for uncountable experiments, it is necessary that we choose a σ-field that is smaller than the power set.

Example 2.23. From Example 2.19, we know that $\{\phi, \Omega\}$ is a field; it obviously satisfies properties (i)-(ii). Of course, this trivial field contains no useful information about any particular experiment. Consider the experiment of tossing a single die with outcomes $\{1, 2, 3, 4, 5, 6\}$. The power set $\mathcal{P}(\Omega)$ consists of $2^6 = 64$ elements. However, it is possible to define a smaller (nontrivial) field, though as we shall see it is not as useful as the power set. Consider the following two events: $E = \{1, 3, 5\}$ and $F = \{2, 4, 6\}$, and observe that $\mathcal{F} = \{\phi, \Omega, E, F\}$ is a field. Properties (i) and (ii) are satisfied; in particular, $E = F^c \in \mathcal{F}$ and $E \cup F = \Omega \in \mathcal{F}$. Although this is a valid field, it results in a simplification of the experiment such that we are concerned only about the outcome being odd or even; the specific number showing on the die is of no interest. In fact, by defining such a field, this "simplified" experiment is structurally identical to a single toss of a coin where, for example, $H \equiv$ even number and $T \equiv$ odd number showing on the die. Tossing a single die and using the smaller field above *simulates* tossing a single coin.

Consider again the situation where we start with some subset of Ω and then generate a field that includes the subset.

Example 2.24. For the experiment of tossing a single die, we would like to generate a field starting with the subset $\{1, 2\}$. Observe that $\{1, 2\}^c = \{3, 4, 5, 6\}$, $\{1, 2\} \cap \{3, 4, 5, 6\} = \phi$, and $\{1, 2\} \cup \{3, 4, 5, 6\} = \Omega$. Thus the field *generated* by this subset is $\mathcal{F} = \{\phi, \Omega, \{1, 2\}, \{3, 4, 5, 6\}\}$. Obviously, this field fails to capture all outcomes and events of the original experiment; in fact, it is equivalent to a binary experiment such as the toss of a single coin (as in Example 2.23). Consider starting with the subset $\{\{1, 2\}, \{3, 4\}\}$. Then it is easy to show that the generated field is $\mathcal{F} = \{\phi, \Omega, \{1, 2\}, \{3, 4\}, \{5, 6\}, \{3, 4, 5, 6\}, \{1, 2, 5, 6\}, \{1, 2, 3, 4\}\}$. In order to capture all details of the single-die experiment, we would need to choose the power set for the field with $2^6 = 64$ elements.

These last examples illustrate the difference between a "fine" field (the power set) and a "coarse" field (given by some subset of the power set). Later when conditional probability is covered, we use this distinction between fine and coarse fields to derive a useful property in probability theory when conditioning on events.

Next, we describe a property of a field $\mathcal{F}$ when Ω is finite, based on the definition of atoms.

Definition: Atoms of a Field The smallest events in $\mathcal{F}$ (excluding ϕ) are *atoms*. We denote the set of atoms by $\mathcal{A}$.

This definition means that an atom cannot be obtained from the union of other events in $\mathcal{F}$, and they are disjoint. Every other event in $\mathcal{F}$ is obtained by set operations on the elements of $\mathcal{A}$ (in order to satisfy the requirements

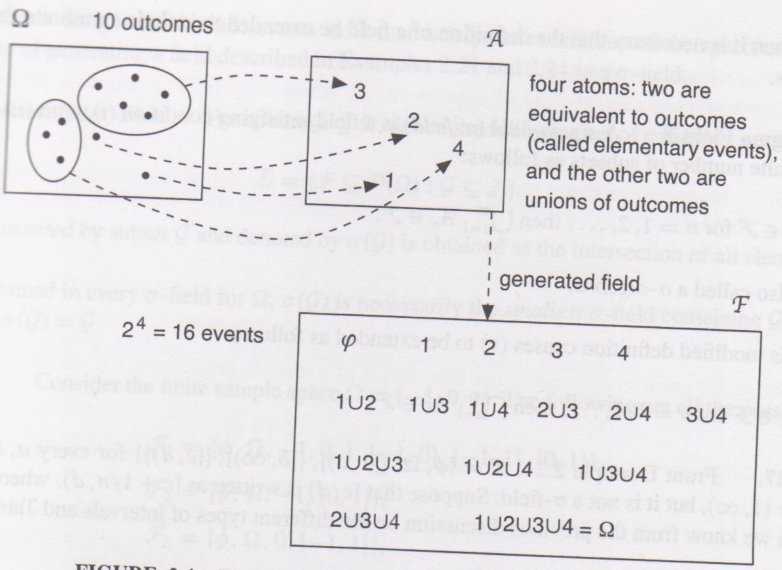

FIGURE 2.4 Example outcomes in Ω, atoms in $\mathcal{A}$, and events in $\mathcal{F}$.

field). A simple finite example showing the connection between outcomes in Ω, atoms in $\mathcal{A}$, and events in $\mathcal{F}$ is illustrated in Figure 2.4.

Example 2.25. For the second field in Example 2.24, the three atoms are $\mathcal{A} = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$, which are not outcomes in Ω.

It is clear that the atoms of the power set for an experiment are the elementary events (outcomes). If individual outcomes are excluded from the field, then the atoms are nonelementary events as shown in Example 2.24.

Theorem 2.1. The number of elements in any finite field $\mathcal{F}$ is given by 2^n, where $n = |\mathcal{A}|$.

Proof. From the definition of an atom and the requirements of a field, the elements of $\mathcal{F}$ must be the power set $\mathcal{P}(\mathcal{A})$ of the atoms. We have already shown that the cardinality of the power set is 2^N for a set of size N, and thus $N = n$ completes the proof. □

This theorem is useful because we can immediately determine that $\mathcal{F}$ is *not* a field if the number of elements is not 2, 4, 8, 16, and so on. Note that Ω need not be discrete in order for the field to be finite, as demonstrated in Example 2.26.

Example 2.26. Let $\Omega = [0, \infty) = \mathcal{R}^+$ and $\mathcal{F} = \{\phi, \Omega, [0, c), [c, \infty)\}$ for some $c > 0$. Thus, although the sample space is continuous, the field has only four elements with atoms given by two subintervals: $[0, c)$ and $[c, \infty)$. Suppose instead that the atoms are $[0, c)$, $[c, d)$, and $[d, \infty)$ for some $d > c > 0$. From the theorem above, we know for this continuous sample space that the field has eight elements: $\mathcal{F} = \{\phi, \Omega, [0, c), [c, d), [d, \infty), [0, d), [c, \infty), [0, c) \cup [d, \infty)\}$. Such intervals on $\mathcal{R}$ are important for continuous sample spaces, as discussed later when we introduce Borel sets.

For problems involving a finite sample space, the definition of a field with conditions (i) and (ii) is adequate for us to specify a probability space. However, if the sample space is infinite, either countably infinite or uncountable, then it is necessary that the definition of a field be extended to include an infinite (but countable) union of events.

Definition: Sigma Field $\mathcal{F}$ A *sigma-field* (σ*-field*) is a field satisfying condition (i), with (ii) extended to a countably infinite number of subsets as follows:

(ii′) If $A_n \in \mathcal{F}$ for $n = 1, 2, \ldots$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

A σ-field is also called a σ*-algebra*.

Of course, this modified definition causes (v) to be extended as follows:

(v′) If $A_n \in \mathcal{F}$ for $n = 1, 2, \ldots$, then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

Example 2.27. From Example 2.21, $\mathcal{F} = \{\phi, \Omega, \{[1, a)\}, \{[b, \infty)\}, \{[c, d)\}\}$ for every $a, b, c, d > 1$ is a field for $\Omega = [1, \infty)$, but it is not a σ-field. Suppose that $[c, d)$ is written as $[c + 1/n, d)$, where n is a positive integer. Then we know from the previous discussion on the different types of intervals and Table 2.2 that

$$\bigcup_{n=1}^{\infty} [c + 1/n, d) = (c, d), \tag{2.18}$$

which is a type of interval not in $\mathcal{F}$.

Next, we consider the cardinality of infinite sets.

Definition: Beth Numbers The *beth numbers* $\{\beth_n\}$ (Forster, 1995) are used to represent the cardinality of different types of infinite sets.

In mathematics, the cardinality of the natural numbers $\mathcal{N}$ is denoted by the symbol $\beth_0$ which is called *beth null*. The beth numbers are generated by the following recursion:

$$\beth_{n+1} = 2^{\beth_n}, \tag{2.19}$$

which we see has a form similar to that in Proposition 2.1 for the cardinality of $\mathcal{P}$ for a finite set. It turns out that the cardinality of $\mathcal{R}$ is beth one $\beth_1 = 2^{\beth_0}$ which is the size of the power set for $\mathcal{N}$. Similarly, the cardinality of $\mathcal{P}(\mathcal{R})$ is beth two $\beth_2$. We are not concerned with the theory behind the cardinality of infinite sets and the nature of the different types of infinity; instead, we use beth numbers as a convenient notation to represent infinite sets that have the same cardinality. Several examples are provided in Table 2.4, some of which are defined later.

For $\Omega = \mathcal{N}$, the power set $\mathcal{P}(\mathcal{N})$ is a σ-field. In fact, the power set for *any* Ω is a σ-field, and it is clear from the previous discussions that σ-field $\mathcal{F}$ for any Ω satisfies $\{\Omega, \phi\} \subseteq \mathcal{F} \subseteq \mathcal{P}(\Omega)$.

Definition: Event Space An *event space* is the set of events in Ω that comprise the σ-field $\mathcal{F}$. The event space is written as the double $\{\Omega, \mathcal{F}\}$.

TABLE 2.4 Beth Numbers and Cardinality of Infinite Sets

Beth Number	Infinite Sets with the Same Cardinality
$\beth_0$	$\mathcal{N}, \mathcal{Q}, \mathcal{Z}$
$\beth_1$	$\mathcal{R}, \mathcal{R} - \mathcal{Q}, \mathcal{P}(\mathcal{N}), \mathcal{B}(\mathcal{R}), \mathcal{C}, [a, b] \subset \mathcal{R}$, Cantor set
$\beth_2$	$\mathcal{P}(\mathcal{R}), \mathcal{P}(\mathcal{P}(\mathcal{N}))$

The event space is designed to include events in Ω that satisfy the conditions of the specific σ-field $\mathcal{F}$. We extend the idea of generating a field described in Examples 2.21 and 2.24 to a σ-field.

Definition: Generated σ-Field For subset $\mathcal{G} \subseteq \Omega$, define the following set of σ-fields:

$$E = \{\mathcal{F} \subseteq \mathcal{P}(\Omega) : \mathcal{G} \subseteq \mathcal{F}\}. \tag{2.20}$$

The σ-field *generated* by subset $\mathcal{G}$ and denoted by $\sigma(\mathcal{G})$ is obtained as the intersection of all elements in E.

Since $\mathcal{G}$ is contained in every σ-field for Ω, $\sigma(\mathcal{G})$ is necessarily the *smallest* σ-field containing $\mathcal{G}$. If $\mathcal{G}$ itself is a σ-field, then $\sigma(\mathcal{G}) = \mathcal{G}$.

Example 2.28. Consider the finite sample space $\Omega = \{-1, 0, 1\}$. The following are all the possible σ-fields:

$$\mathcal{F}_1 = \{\phi, \Omega, -1, 0, 1, \{-1, 0\}, \{-1, 1\}, \{0, 1\}\}, \tag{2.21}$$

$$\mathcal{F}_2 = \{\phi, \Omega, -1, \{0, 1\}\}, \tag{2.22}$$

$$\mathcal{F}_3 = \{\phi, \Omega, 0, \{-1, 1\}\}, \tag{2.23}$$

$$\mathcal{F}_4 = \{\phi, \Omega, 1, \{-1, 0\}\}, \tag{2.24}$$

$$\mathcal{F}_5 = \{\phi, \Omega\}. \tag{2.25}$$

If $\mathcal{G} = \Omega$, then $\sigma(\mathcal{G}) = \mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{F}_3 \cap \mathcal{F}_4 \cap \mathcal{F}_5 = \mathcal{F}_5$ which is the trivial σ-field. If $\mathcal{G} = \{1\}$, then $\sigma(\mathcal{G}) = \mathcal{F}_1 \cap \mathcal{F}_4 = \mathcal{F}_4$. If $\mathcal{G} = \{0, 1\}$, then $\sigma(\mathcal{G}) = \mathcal{F}_1$, which is the power set $\mathcal{P}(\Omega)$.

We are interested in σ-fields that are *measurable* so that probabilities can be assigned in a consistent manner to the event space $\{\Omega, \mathcal{F}\}$. For finite and countably infinite experiments, there is generally no need to consider σ-fields smaller than the power set $\mathcal{P}(\Omega)$. However, for continuous sample spaces, the power set is "too large." The Vitali set described at the end of this chapter for $\Omega = [0, 1]$ is an example of a subset that is not measurable. It will be convenient from a practical sense, as well as for most applications, to consider *intervals* for continuous experiments. Also, since sample spaces are mapped to the real line $\mathcal{R}$ for the random variables in Chapter 3, we are interested in a specific type of σ-field for intervals on $\mathcal{R}$ called the Borel σ-field.

Definition: Borel σ-Field The *Borel σ-field* $\mathcal{B}(\mathcal{R})$ on the real line is the smallest σ-field generated by all open intervals in $\Omega = \mathcal{R}$. Elements of $\mathcal{B}(\mathcal{R})$ are called *Borel sets.*

The Borel σ-field is generated by starting with all open intervals in $\mathcal{R}$ of the form (a, b). Individual points (singletons) are included by observing that

$$a = \lim_{n\to\infty} (a - 1/n, a + 1/n). \tag{2.26}$$

This result allows us to include all semi-open intervals:

$$(a, b] = (a, b) \cup b, \qquad [a, b) = (a, b) \cup a. \tag{2.27}$$

Since $(-\infty, b] = (b, \infty)^c$ and $[a, \infty) = (-\infty, a)^c$, all closed intervals are also included:

$$[a, b] = [a, \infty) \cap (-\infty, b]. \tag{2.28}$$

Thus, $\mathcal{B}(\mathcal{R})$ consists of all types of intervals on $\mathcal{R}$ as well as all individual points. It is not easy to visualize subsets of $\mathcal{R}$ that lie outside of $\mathcal{B}(\mathcal{R})$; we present an example of such a subset when measure theory is discussed later. However, it turns out that these subsets do not arise in practical applications and so they are not of concern. Since we will also consider σ-fields in $\Omega = \mathcal{R}^N$ when *random vectors* are covered in Chapter 4, the above definition extends to the smallest σ-field of subsets of $\mathcal{R}^N$ denoted by $\mathcal{B}(\mathcal{R}^N)$, which is defined to be open *rectangles* on $\mathcal{R}^2$ and open hyper-rectangles on $\mathcal{R}^N$ (for $N > 2$).

Example 2.29. Examples of *Borel sets* include the following: (i) closed interval $[a, b]$, (ii) open interval (a, b), (iii) set of rational numbers $\mathcal{Q}$, (iv) set of irrational numbers $\mathcal{R} - \mathcal{Q}$, (v) $\{1, 2, 3, 4, 5, 6\}$, and (vi) the Cantor set described later.

2.5 SUMMARY OF A RANDOM EXPERIMENT

It will be useful at this point to summarize our description of a random experiment for which we want to assign a probability measure:

- The *sample space* Ω is the collection of all outcomes of an experiment. Depending on the type of experiment, it may contain a finite, countably infinite, or uncountable number of elements.
- An *event* of an experiment is a subset of Ω consisting of one or more outcomes, and which usually share some feature. Since the goal is to assign a probability measure to events that is consistent, we may not want to include all possible subsets of Ω.
- A *σ-field* $\mathcal{F}$ is a collection of subsets of Ω that satisfy the algebraic conditions in (i) and (ii′), as well as (iii), (iv), and (v′). One can envision constructing a σ-field by starting with some subset of Ω and expanding the number of subsets until the collection satisfies (i) and (ii′).
- If Ω is finite or countably infinite, then all subsets described by the power set $\mathcal{P}(\Omega)$ comprise a useful σ-field $\mathcal{F}$ for the experiment.
- If Ω is uncountable, then the Borel σ-field $\mathcal{B}(\mathcal{R})$ is a useful σ-field for the real line that consists of all intervals (open, semi-open, and closed) as well as individual points.

With these definitions of the sample space Ω and the event space $\{\Omega, \mathcal{F}\}$, we have a collection of events that satisfy the additivity property in (ii′). It is now possible to formulate a *probability space* which we denote by the triple $\{\Omega, \mathcal{F}, P\}$. We begin with a brief discussion of measure theory, of which the probability measure is a particular case.

2.6 MEASURE THEORY

Measure theory is concerned with assigning a "size" called a *measure* to subsets of a sample space Ω. Depending on the problem, this size might be a length, an area, or a volume in Euclidean space; this is known as the Lebesgue measure which is defined below. In our case, we are interested in assigning a *probability measure* to subsets of Ω that comprise the σ-field $\mathcal{F}$. For uncountable experiments, it is not possible to assign probabilities in a consistent manner to $\mathcal{P}(\mathcal{R})$. The power set of $\Omega = \mathcal{R}$ is too large (it has cardinality beth two $\beth_2$), so instead we choose $\mathcal{F}$ to be $\mathcal{B}(\mathcal{R})$, which is the Borel σ-field on the real line and has cardinality beth one $\beth_1$ (the same as $\mathcal{R}$). The Borel σ-field is useful for essentially all practical applications because we are generally interested in events described by *intervals* on the real line.

Let μ be the notation for a measure that we want to define for the event space, and denote the *measure space* by the triple $\{\Omega, \mathcal{F}, \mu\}$. (Note that μ should not be confused with the mean of a random variable covered later in Chapter 5.)

Definition: Measure A *measure* is a mapping of events $\{E_n\} \in \mathcal{F}$ to $\mathcal{R}$ that has the following properties:

- $\mu(E_n) \geq 0$ for all events $E_n \in \mathcal{F}$.
- $\mu(\phi) = 0$.

- For all events $\{E_n\}$ such that $E_n \cap E_m = \phi$ for $n \neq m$ (mutually exclusive):

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n). \tag{2.29}$$

Similar properties are shown later for the probability measure where they are known as the *three axioms of probability*.

Example 2.30. The *trivial measure* is $\mu(E_n) = 0$ for all $E_n \in \mathcal{F}$ in the event space $\{\Omega, \mathcal{F}\}$.

Example 2.31. For $\Omega = \mathcal{R}$ and $\mathcal{F} = \mathcal{B}(\mathcal{R})$, the *standard Gaussian measure* is

$$\mu(E) = \frac{1}{\sqrt{2\pi}} \int_E \exp(-x^2/2)dx, \tag{2.30}$$

which turns out to be the probability $P(E)$ of event E for the standard Gaussian random variable X with zero mean and unit variance.

Example 2.32. The *Dirac measure* for event E in the event space $\{\Omega, \mathcal{F}\}$ is given by

$$\delta_\omega(E) \triangleq \begin{cases} 1, & \omega \in E \\ 0, & \text{else.} \end{cases} \tag{2.31}$$

It can be shown that $\delta_\omega(E)$ is a probability measure, and thus E is the almost sure event in Ω (see Problem 2.23).

Consider two important measures that will be useful later.

Definition: Counting Measure Let Ω be a countable sample space and $\mathcal{P}(\Omega)$ its power set. The *counting measure* for any set $E \in \mathcal{P}(\Omega)$ is

$$C(E) = \begin{cases} |E|, & E \text{ is a finite subset} \\ \infty, & \text{otherwise,} \end{cases} \tag{2.32}$$

where $|E|$ is the cardinality of E. The infinity in this definition is actually beth null $\beth_0$ mentioned above, the size of a countably infinite set.

Obviously, this measure is useful only for finite sets; it certainly provides no information about the size of uncountable events, such as intervals on $\mathcal{R}$. For this latter case, we consider the Lebesgue measure.

Definition: Lebesgue Measure on $\mathcal{R}$ The *Lebesgue measure* of interval $[a, b] \in \mathcal{B}(\mathcal{R})$ is

$$L([a, b]) \triangleq b - a, \tag{2.33}$$

which is the *length* of the interval. Since the Lebesgue measure of a single point (singleton) is defined to be zero, we also have $L((a, b)) = L((a, b]) = L([a, b)) = b - a$.

In order to extend the Lebesgue measure to N-dimensional Euclidean space, it will be convenient to define the Cartesian product.

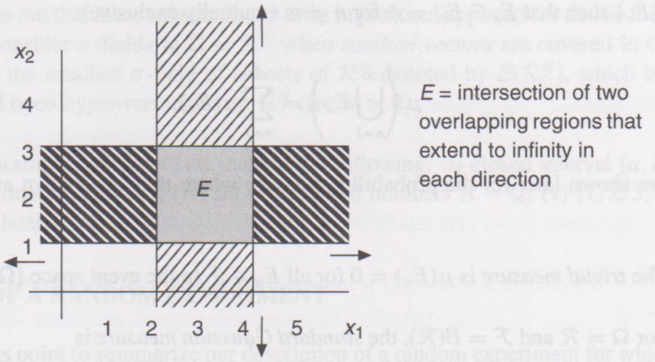

FIGURE 2.5 Cartesian product $E = [2, 4] \times [1, 3]$ for Example 2.33.

Definition: Cartesian Product The *Cartesian product* of a set of intervals $\{[a_n, b_n]\}$, $n = 1, \ldots, N$, is

$$E \triangleq \{(x_1, \ldots, x_N) : x_1 \in [a_1, b_1], \ldots, x_N \in [a_N, b_N]\}. \tag{2.34}$$

E corresponds to an ordered N-tuple, which can be represented using the following product notation:

$$E = [a_1, b_1] \times \cdots \times [a_N, b_N] = \prod_{n=1}^{N} [a_n, b_n]. \tag{2.35}$$

The Cartesian product of intervals on $\mathcal{R}$ is a rectangle on $\mathcal{R}^2$ and a hyper-rectangle on $\mathcal{R}^N$ (for $N > 2$).

Example 2.33. For intervals $[1, 3]$ and $[2, 4]$, the Cartesian product $E = \{(x_1, x_2) : x_1 \in [2, 4], x_2 \in [1, 3]\}$ is a rectangle on $\mathcal{R}^2$ described by all points on the abscissa in the interval $[2, 4]$ *intersected* with all points on the ordinate in the interval $[1, 3]$. This is depicted in Figure 2.5.

For the real line $\mathcal{R}$, we can write $\mathcal{R}^2 = \mathcal{R} \times \mathcal{R}$ (a plane), $\mathcal{R}^3 = \mathcal{R}^2 \times \mathcal{R} = \mathcal{R} \times \mathcal{R} \times \mathcal{R}$ (a hyperplane), and so on. With this geometric viewpoint, we readily see in general that the Lebesgue measure of E corresponds to the area of a rectangle, and in three dimensions it is a *volume*. For the N-dimensions, we have the following definition.

Definition: Lebesgue Measure on $\mathcal{R}^N$ The *Lebesgue measure* of the Cartesian product $E = \prod_{n=1}^{N} [a_n, b_n] \in \mathcal{B}(\mathcal{R}^N)$ is

$$L(E) \triangleq \prod_{n=1}^{N} (b_n - a_n), \tag{2.36}$$

which is the hyper-volume of the corresponding hyper-rectangle on $\mathcal{R}^N$.

Example 2.34. For the Cartesian product in Example 2.33, $L([1, 3] \times [2, 4]) = 4$ is the area of a rectangle, and $L([8, 10] \times [1, 2] \times [4, 8]) = 8$ is the volume of a rectangular cuboid.

The Cantor set described in Example 2.35 is interesting because of the way it is constructed, and because it has Lebesgue measure zero.